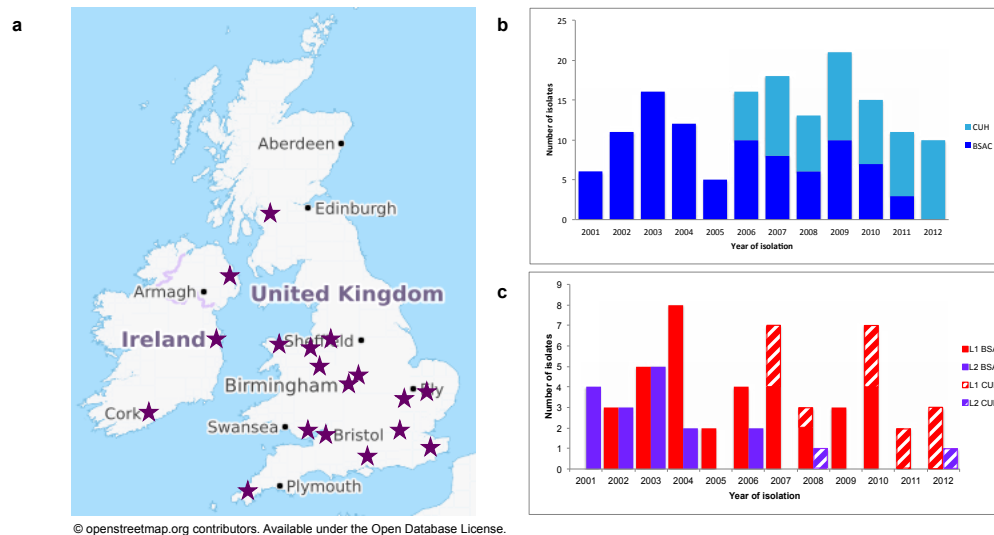


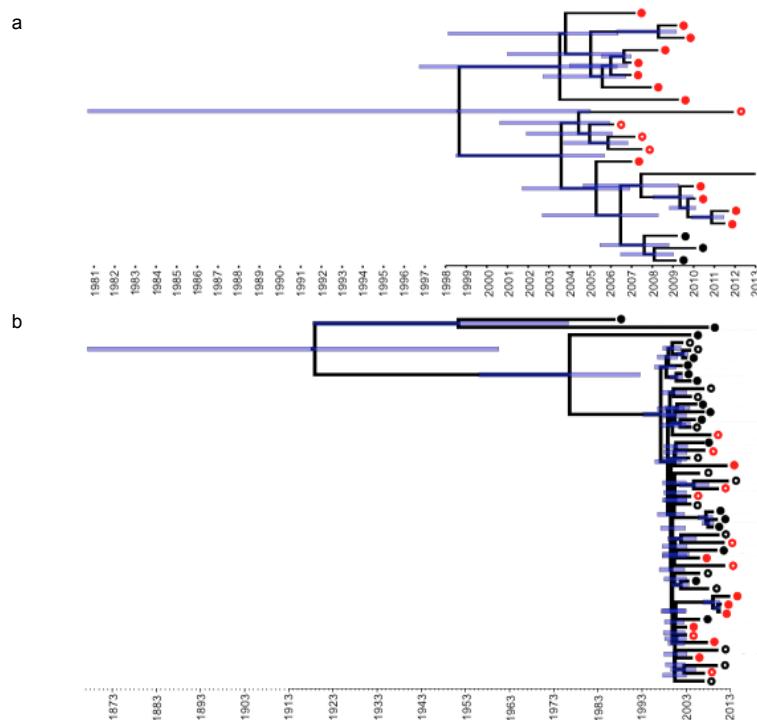
**Genome-based characterization of hospital-adapted *Enterococcus faecalis* lineages**

Kathy E. Raven, Sandra Reuter, Theodore Gouliouris, Rosy Reynolds, Julie E Russell, Nicholas M. Brown, M. Estée Török, Julian Parkhill, Sharon J. Peacock.



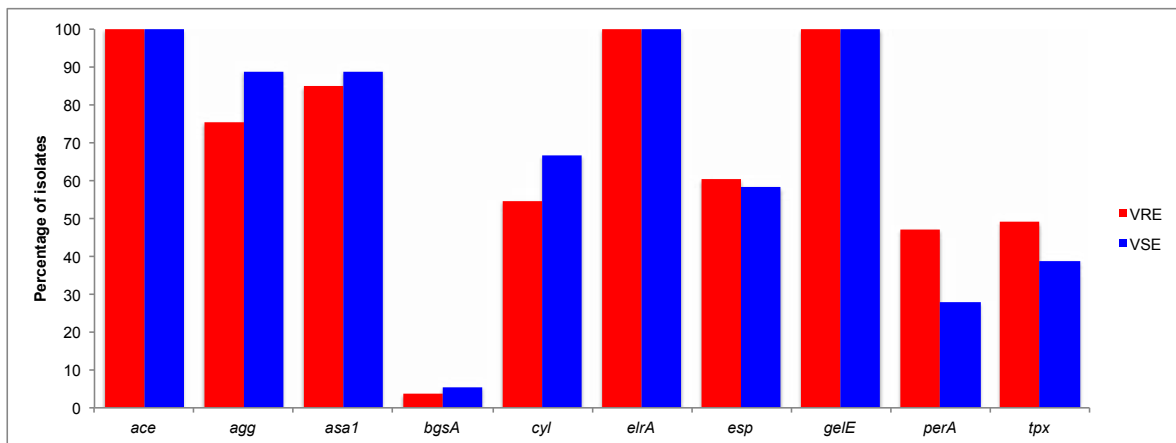
## Supplementary Figure 1

Geographical and temporal spread of the study isolates from the British Society for Antimicrobial Chemotherapy (BSAC) and Cambridge University Hospitals NHS Foundation Trust (CUH). (a) BSAC study isolates were from the contributing laboratories indicated by a purple star. (b) Year of isolation for the BSAC and CUH isolates. (c) Year of isolation for L1 and L2.



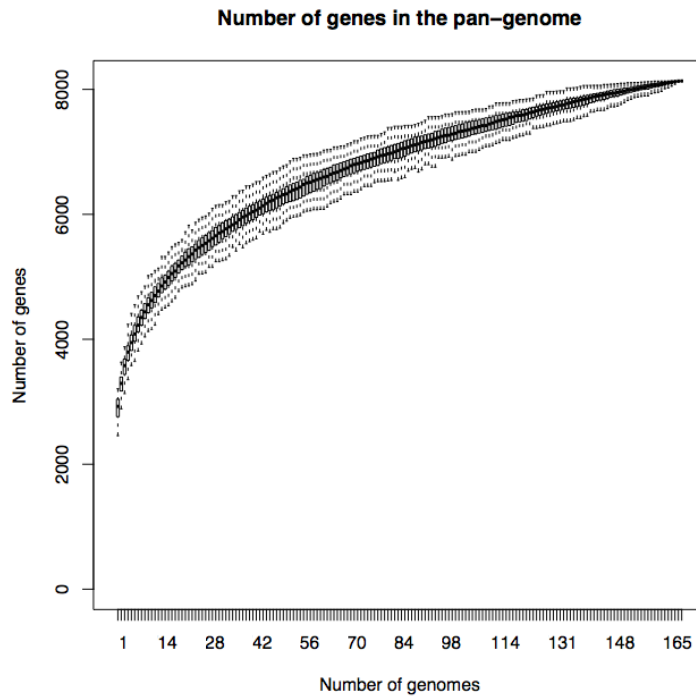
### Supplementary Figure 2

Timeline of the emergence of two dominant *E. faecalis* lineages. Maximum clade credibility tree for L3 (a) and L1 (b), based on median values from BEAST analyses. The time scale is shown along the bottom. Red circles represent isolates from CUH, and black dots represent isolates from the East Midlands referral network (a), or elsewhere (b). Closed circles denote VRE and open circles denote VSE. Blue bars indicate 95% highest probability density intervals.



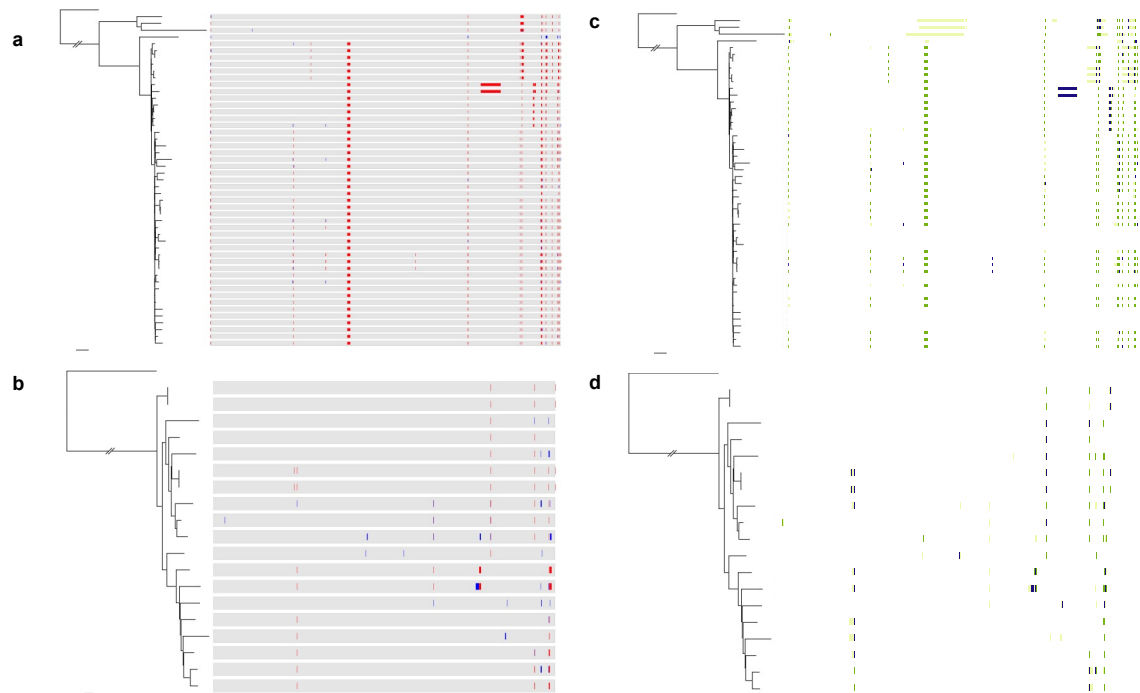
### Supplementary Figure 3

Prevalence of candidate virulence genes in VREs and VSEs. Graph comparing the percentage of VREs (n=53) and VSEs (n=36) isolates from the three dominant lineages (L1-3) in which candidate virulence genes were detected. Of the eighteen candidate genes investigated, eight were ubiquitous in the study collection and were not included in this figure.



#### Supplementary Figure 4

Pangenome of *E. faecalis*. Cumulative number of genes for 168 *E. faecalis* isolates. The analysis was run 100 times and the error bars showing the standard deviation are indicated.



### Supplementary Figure 5

Recombination within L1 (a,c) and L2 (b,d). Maximum likelihood trees based on SNPs in the core genome after recombination events (identified by Gubbins) were removed, rooted on an outlier. (a) and (b): the colored blocks on the grey lines represent regions of recombination identified by Gubbins: red indicates a recombination event that occurred ancestral to the individual isolate and blue indicates that the recombination event is specific to that isolate. (c) and (d) show recombination events identified by Gubbins (green), BratNextGen (lime) or both (blue). Scale bars indicate 25 SNPs.

**Supplementary Table 2**

Accessory genes unique to the dominant lineages and present in more than 10 isolates.

Number of genes with this annotation	Annotation	Number of isolates
60	hypothetical, unknown or uncharacterized protein	Each of the genes in this category were variable present in the population, (range 11-49 isolates)
5	Cro/CI family transcriptional regulator	Each of the genes in this category were variable present in the population, (range 17-48 isolates)
1	Uncharacterized phage-associated protein	48
1	phage integrase family site specific recombinase	46
1	phage integrase family site specific recombinase	21
1	phage DNA methylase	15
1	phage terminase-like protein large subunit	15
1	phage putative tail component protein	15
1	phage major tail protein, phi13 family	15
1	phage portal protein, HK97 family	15
1	phage prohead protease, HK97 family	15
1	putative phage head-tail adaptor	15
1	major capsid protein	15
1	phage tail tape measure protein, TP901 family, core region	15
1	phage terminase, large subunit, PBSX family	14
1	phage terminase domain A protein	14
1	Phage terminase, small subunit	12
1	transposase	21
1	transposase family protein	21
1	IS4 family transposase	19
1	IS6 family transposase	13
1	WxL domain surface protein	49
1	helix-turn-helix domain-containing protein	46
1	ATP-binding protein	46
1	transcriptional regulator, y4mF family	45
1	choline binding protein	38
1	putative lipoprotein	23
1	structural protein	22
1	conjugative transposon membrane protein	21
1	Host cell surface-exposed lipoprotein.	21
1	SOS-response repressor and protease LexA	21
1	HNH endonuclease domain protein	21
1	Helix-turn-helix domain	21
1	recombination protein F	19
1	Uncharacterized distant relative of cell wall-associated hydrolases	19
1	ski2-like helicase	19
1	collagen adhesion CnaB-type domain protein	19
1	DNA repair protein RadC	19
1	Restriction endonuclease S subunits	18
1	VirE domain protein	18
1	Integrase core domain protein	17
1	plasmid recombination enzyme Mob	15
1	HeH/LEM domain	15
1	HNH endonuclease domain protein	15
1	BRO family transcriptional regulator	13
1	drug resistance transporter, putative	11
1	aspartate 1-decarboxylase domain-containing protein	11

**Supplementary Table 3**

Non-synonymous SNPs unique to L3.

Location in V583 genome	Annotation
23537	PTS system, mannose-specific IIAB components
92310	diacylglycerol kinase catalytic domain protein
95588	cell wall surface anchor family protein
107113	transcriptional regulator, ArgR family
126919	conserved hypothetical protein
198651	ribosomal protein s10
244870	chaperonin, 33 kDa
257798	PTS system, beta-glucoside-specific IIABC component
258122	PTS system, beta-glucoside-specific IIABC component
341571	sensor histidine kinase
379256	transcriptional regulator, putative
443918	ferrous iron transport protein B
474028	thermonuclease precursor
474053	thermonuclease precursor
474200	thermonuclease precursor
474313	thermonuclease precursor
474508	thermonuclease precursor
474512	thermonuclease precursor
474513	thermonuclease precursor
474514	thermonuclease precursor
474516	thermonuclease precursor
474518	thermonuclease precursor
545546	ABC transporter, ATP-binding protein/permease protein
618410	conserved hypothetical protein
635256	conserved hypothetical protein
638033	conserved hypothetical protein
669810	trigger factor
673177	PTS system, fructose-specific family, IIABC components
693889	carbamate kinase
704472	rhodanese family protein
708208	hypothetical protein
719566	excinuclease ABC, subunit B
765966	amino acid ABC transporter, ATP-binding protein
847297	primosomal protein DnaI
860856	glycerol dehydrogenase, putative
863075	preprotein translocase, YajC subunit, putative
878217	peptide ABC transporter, ATP-binding protein
880999	translation initiation factor IF-3
896214	hypothetical protein
917473	aldose 1-epimerase, putative



938230	exodeoxyribonuclease VII, large subunit
955856	cell division protein FtsZ
957653	conserved hypothetical protein
964477	glucose-6-phosphate 1-dehydrogenase
978765	glycosyl hydrolase, family 1
979862	conserved hypothetical protein
988543	hydrolase, alpha/beta hydrolase fold family
995113	6-aminohexanoate-cyclic-dimer hydrolase, putative
1046518	acetyltransferase, GNAT family
1082069	exonuclease RexB
1085230	exonuclease RexA
1129958	N4-(beta-N-acetylglucosaminy)-L-asparagine, putative
1221577	conserved hypothetical protein
1286758	conserved domain protein
1288458	permease domain protein
1335097	dihydroxyacetone kinase family protein
1340299	acetyl-CoA acetyltransferase/hydroxymethylglutaryl-CoA reductase, degradative
1340394	acetyl-CoA acetyltransferase/hydroxymethylglutaryl-CoA reductase, degradative
1366698	NAD-dependent formate dehydrogenase, beta subunit, putative
1366947	NAD-dependent formate dehydrogenase, beta subunit, putative
1388715	conserved hypothetical protein
1395937	glutamate dehydrogenase
1470141	transcription antiterminator, bglG family
1497784	ferredoxin
1521341	prephenate dehydratase
1573410	DNA topoisomerase IV, A subunit
1598884	heat shock protein, HslVU, ATPase subunit HslU
1663957	carbamoyl-phosphate synthase, large subunit
1716571	hypothetical protein
1717396	conserved hypothetical protein
1741118	conserved hypothetical protein
1758290	drug resistance transporter, EmrB/QacA family protein
1858075	ATP-dependent Clp protease, ATP-binding subunit ClpX
1859263	conserved hypothetical protein
1860566	acetyltransferase, GNAT family
1860947	C4-dicarboxylate anaerobic carrier
1877393	conserved hypothetical protein
1877539	conserved hypothetical protein
1877548	conserved hypothetical protein
1877752	conserved hypothetical protein
1877788	conserved hypothetical protein
1878240	cation-transporting ATPase, E1-E2 family
1881168	conserved hypothetical protein
1882000	cyclic AMP receptor protein, putative

1882256	cyclic AMP receptor protein, putative
1882700	hypothetical protein
1887227	conserved hypothetical protein
1889869	PTS system, IID componenet
1890228	PTS system, IID componenet
1892604	sigma-54 dependent DNA-binding response regulator
1893048	sigma-54 dependent DNA-binding response regulator
1893720	sigma-54 dependent DNA-binding response regulator
1893908	sigma-54 dependent DNA-binding response regulator
1908014	aspartyl-tRNA synthetase
1922107	competence protein
1968954	ABC transporter, permease protein, putative
2090904	membrane protein, putative
2310643	conserved hypothetical protein
2363200	2-dehydropantoate 2-reductase, putative
2401439	conserved hypothetical protein
2508419	conserved hypothetical protein
2511350	ABC transporter, ATP-binding protein/permease protein
2567048	alcohol dehydrogenase, zinc-containing
2660959	permease protein, putative
2664330	anaerobic ribonucleaside-triphosphate reductase
2696400	membrane protein, putative
2737944	penicillin bindins protein 2B
2758860	3-oxoacyl- (acyl-carrier-protein) synthase II
2759471	3-oxoacyl- (acyl-carrier-protein) synthase II
2777217	peptidyl-prolyl cis-trans isomerase, cyclophilin-type
2785518	ABC transporter,permease protein
2785745	ABC tranporter, ATP-binding protein
2800512	conserved hypothetical protein
2893920	sodium:dicarboxylate symorter family protein
2967196	glyoxalase family protein
3012753	PTS system, IIC component
3012754	PTS system, IIC component
3025245	conserved hypothetical protein
3040605	DNA mismatch repair protein HexA
3058822	conserved hypothetical protein
3072733	ABC tranporter, ATP-binding protein
3171931	sensor histidine kinase

**Supplementary Table 4**

Insertion sites of the *vanA* transposon in the dominant lineages. Table describing the type, frequency and variations of insertion sites and the matches using BLAST. Colors in the left hand column correspond to those in Figure 4.

	Type	No. isolates	Variation within types	Best enterococcal match	Variations between types
	1A	24	★❖⊙	pIP816 99% cover 98% ID	
	1B	3	★❖	pTW9* 71% cover 99% ID	Identical to 1A for the first 4,270bp
	1C	2		pTEF1 78% cover 100% ID	Identical to 1A for the first 4,798bp
	2A	3		Aus85 p1 89% cover 100% ID	Identical to 2B and 2C for the first 239bp
	2B	1		pF856 22% cover 100% ID	Identical to 2C for the first 2,425bp
	2C	5	⊙	pF856 65% cover 100% ID	
	3	2		pLG2** 84% cover 99% ID	
	<250bp	11	6 different variants		

★ Deletion of 21bp

❖ Insert of 1202bp to 1620bp of sequence

⊙ Change at the end of the available sequence of <100bp

Note: excluded *Enterococcus* sp. 7L76 draft genome as chromosome and plasmids may not be separated.

\* These sequences best matched either *Lactococcus garviae* pKL0018, Shuttle vector pKV12 or *Enterococcus* sp. 7L76 draft genome with an equal ID and equal or lower percent coverage. \*\* This sequence best matched to *Staphylococcus aureus* SA268 with 100% cover and 99% ID.